

## **CLAIMS**

What is claimed is:

1    1.    A computer-implemented method for optimizing allocation of computer resources,  
2    comprising:

3         establishing a server model including one or more server nodes, wherein each  
4    server node has an associated set of capacity attributes;

5         establishing a service model including one or more service nodes, wherein each  
6    service node has an associated set of demand attributes;

7         selecting one of a plurality of user-selectable optimization methods; and

8         generating an optimized mapping of the server nodes in the server model to the  
9    service nodes in the service model using the selected optimization method and demand  
10   and capacity attributes.

1    2.    The method of claim 1, further comprising normalizing the capacity attributes of  
2    server nodes and demand attributes of service nodes.

1    3.    The method of claim 2, further comprising:

2         selecting one of a plurality of user-selectable objective functions, wherein each of  
3    the objective functions evaluates a solution as a function of the demand and capacity  
4    attributes; and

5         generating an optimized mapping of the server nodes to the service nodes using the  
6    selected one of the objective functions and selected one of the optimization methods.

1    4.    The method of claim 3, wherein the plurality of optimization methods include a  
2    genetic process and a complete search process.

1    5.    The method of claim 3, further comprising:

2         establishing one or more service-node relationships between selected pairs of the  
3    service nodes, wherein each service-node relationship has an associated transport demand  
4    attribute specifying a quantity of communication resources required for communication  
5    between the associated pair of service nodes;

6       establishing one or more server-node relationships between selected pairs of the  
7   server nodes, wherein each server-node relationship has an associated transport capacity  
8   attribute specifying a quantity of communication resources available for communication  
9   between the associated pair of server nodes; and  
10      generating the optimized mapping as a function of the service-node relationships  
11   and server-node relationships.

1   6.     The method of claim 4, wherein the plurality of objective functions includes a first  
2   function for quantifying a balance processing load between nodes and a second function  
3   quantifies a transport demand between the nodes.

1   7.     The method of claim 4, further comprising:  
2       establishing a plurality of server models, each server model including one or more  
3   server nodes, wherein each server node has an associated set of capacity attributes;  
4       designating a layered relationship between the server models, wherein for a first  
5   server-model layer immediately above a second server-model layer, the second server-  
6   model layer includes respective models that represent the nodes in the first server-model  
7   layer;  
8       establishing a plurality of service models, each service model including one or  
9   more service nodes, wherein each service node has an associated set of demand attributes;  
10   and  
11       designating a layered relationship between the service models, wherein for a first  
12   service-model layer immediately above a second service-model layer, the second service-  
13   model layer includes respective models that represent the nodes in the first server-model  
14   layer;  
15       normalizing the capacity attributes of server nodes and demand attributes of  
16   service nodes of the server models and the service models, respectively; and  
17       generating an optimized mapping of the server nodes in a user-selected one of the  
18   server models to service nodes in a user-selected one of the service models using the  
19   selected optimization method.

1       8.     The method of claim 7, further comprising:  
2             establishing one or more service-node relationships between selected pairs of the  
3     service nodes, wherein each service-node relationship has an associated transport demand  
4     attribute specifying a quantity of communication resources required for communication  
5     between the associated pair of service nodes;

6             establishing one or more server-node relationships between selected pairs of the  
7     server nodes, wherein each server-node relationship has an associated transport capacity  
8     attribute specifying a quantity of communication resources available for communication  
9     between the associated pair of server nodes; and

10             generating the optimized mapping as a function of the service-node relationships  
11     and server-node relationships.

1       9.     The method of claim 7, wherein each service node has an associated set of capacity  
2     attributes and further comprising generating an optimized mapping of service nodes in a  
3     first user-selected service model to service nodes in a second user-selected service model  
4     as a function of the demand attributes of the first service model and capacity attributes of  
5     the second service model.

1       10.    The method of claim 7, wherein each server node has an associated set of demand  
2     attributes and further comprising generating an optimized mapping of server nodes in a  
3     first user-selected server model to server nodes in a second user-selected server model as a  
4     function of the demand attributes of the first server model and capacity attributes of the  
5     second server model.

1       11.    The method of claim 1, further comprising:  
2             selecting one of a plurality of user-selectable objective functions, wherein each of  
3     the objective functions evaluates a solution as a function of the demand and capacity  
4     attributes; and  
5             generating an optimized mapping of the server nodes to the service nodes using the  
6     selected one of the objective functions and selected one of the optimization methods.

1    12.    The method of claim 1, wherein the plurality of optimization methods includes a  
2    genetic process and a complete search process.

1    13.    The method of claim 1, further comprising:  
2         establishing a plurality of server models, each server model including one or more  
3         server nodes, wherein each server node has an associated set of capacity attributes;  
4         designating a layered relationship between the server models, wherein for a first  
5         server-model layer immediately above a second server-model layer, the second server-  
6         model layer includes respective models that represent the nodes in the first server-model  
7         layer;  
8         establishing a plurality of service models, each service model including one or  
9         more service nodes, wherein each service node has an associated set of demand attributes;  
10        and  
11         designating a layered relationship between the service models, wherein for a first  
12         service-model layer immediately above a second service-model layer, the second service-  
13         model layer includes respective models that represent the nodes in the first server-model  
14         layer;  
15         normalizing the capacity attributes of server nodes and demand attributes of  
16         service nodes of the server models and the service models, respectively; and  
17         generating an optimized mapping of the server nodes in a user-selected one of the  
18         server models to service nodes in a user-selected one of the service models using the  
19         selected optimization method.

1    14.    An apparatus for optimizing allocation of computer resources, comprising:  
2         means for establishing a server model including one or more server nodes, wherein  
3         each server node has an associated set of capacity attributes;  
4         means for establishing a service model including one or more service nodes,  
5         wherein each service node has an associated set of demand attributes;  
6         means for selecting one of a plurality of user-selectable optimization methods; and  
7         means for generating an optimized mapping of the server nodes in the server model  
8         to the service nodes in the service model using the selected optimization method and  
9         demand and capacity attributes.

1       15. A system for identifying optimal allocations of computing resources in a data  
2 processing arrangement having a plurality of computing machines that host a plurality of  
3 application processes, comprising:

4             a model repository including a plurality of server models and a plurality of service  
5 models, each server model including one or more server nodes and each server node  
6 having an associated set of normalized capacity attributes, each service model including  
7 one or more service nodes and each service node having an associated set of normalized  
8 demand attributes, wherein the server models are defined in a layered relationship and for  
9 a first server-model layer immediately above a second server-model layer, the second  
10 server-model layer includes respective models that represent the nodes in the first server-  
11 model layer, and the service models are defined in a layered relationship and for a first  
12 service-model layer immediately above a second service-model layer, the second service-  
13 model layer includes respective models that represent the nodes in the first service-model  
14 layer; and

15             an optimization engine coupled to the model repository, the optimization engine .  
16 including a plurality of user-selectable objective functions and a plurality of user-  
17 selectable optimization methods, wherein each of the objective functions evaluates a  
18 mapping as a function of the demand and capacity attributes, and each of the optimization  
19 methods generates mappings of service nodes in a user-selected service model to server  
20 nodes in a user-selected server model and selects an optimal one of the mappings.

1       16. The system of claim 15, further comprising:

2             wherein the model repository further includes one or more service-node  
3 relationships between selected pairs of the service nodes, each service-node relationship  
4 having an associated transport demand attribute that specifies a quantity of communication  
5 resources required for communication between the associated pair of service nodes;

6             wherein the model repository further includes one or more server-node  
7 relationships between selected pairs of the server nodes, each server-node relationship  
8 having an associated transport capacity attribute that specifies a quantity of  
9 communication resources available for communication between the associated pair of  
10 server nodes; and

11           the optimization engine is further configured to generate the optimized mapping as  
12       a function of the service-node relationships and server-node relationships.

1     17.   The system of claim 15, wherein each service node has an associated set of  
2     capacity attributes and the optimization engine is further configured to generate an  
3     optimized mapping of service nodes in a first user-selected service model to service nodes  
4     in a second user-selected service model as a function of the demand attributes of the first  
5     service model and capacity attributes of the second service model.

1     18.   The system of claim 15, wherein each server node has an associated set of demand  
2     attributes and the optimization engine is further configured to generate an optimized  
3     mapping of server nodes in a first user-selected server model to server nodes in a second  
4     user-selected server model as a function of the demand attributes of the first server model  
5     and capacity attributes of the second server model.